An optimal oscillatory phase for pattern reactivation during memory retrieval

Kerren, Casper; Linde Domingo, Juan; Hanslmayr, Simon; Wimber, Maria

DOI:
10.1016/j.cub.2018.08.065

Document Version
Peer reviewed version

Citation for published version (Harvard):
Kerren, C, Linde Domingo, J, Hanslmayr, S & Wimber, M 2018, 'An optimal oscillatory phase for pattern reactivation during memory retrieval', *Current Biology*, vol. 28, no. 21, pp. 3383–3392.e6. https://doi.org/10.1016/j.cub.2018.08.065

An optimal oscillatory phase for pattern reactivation during memory retrieval

Casper Kerrén[1,2*], Juan Linde-Domingo[1*], Simon Hanslmayr[1] & Maria Wimber[1]

[1] School of Psychology, College of Life and Environmental Sciences, University of Birmingham, B15 2TT, UK

[2] Lead Contact

[*] These authors contributed equally

[*] Correspondence: c.kerren@pgr.bham.ac.uk

SUMMARY

Computational models and in vivo studies in rodents suggest that the hippocampal system oscillates between states optimal for encoding and states optimal for retrieval. We here show that in humans, neural signatures of memory reactivation are modulated by the phase of a theta oscillation. EEG was recorded while participants were cued to recall previously learned word-object associations, and time-resolved pattern classifiers were trained to detect neural reactivation of the target objects. Classifier fidelity rhythmically fluctuated at 7-8Hz, and was modulated by theta phase across the entire recall period. The phase of optimal classification was shifted approximately 180° between encoding and retrieval. Inspired by animal work, we then computed "classifier-locked averages" to analyse how ongoing theta oscillations behaved around the time points at which the classifier indicated memory retrieval. We found strong theta (7-8Hz) phase consistency approximately 300ms before the time points of maximal neural memory reactivation. Our findings provide important evidence that the neural signatures of memory retrieval fluctuate and are time-locked to the phase of an ongoing theta oscillation.

KEYWORDS

Oscillations, episodic memory, hippocampus, theta oscillations, phase coding

Introduction

Our episodic memory defines us by storing a record of our past experiences and allowing us to consciously access these records. It is widely agreed that the hippocampus and neocortical areas work in conjunction during the formation and later retrieval of a memory [1-4]. At encoding, the hippocampus is thought to continuously store a sparse and non-overlapping index that points to ongoing activity patterns in cortical space. This hippocampal index can later be reactivated by a reminder, and lead to the reconstruction of a previously stored memory pattern in neocortex [1, 2, 5-8]. Many recent studies have tested these computational assumptions by tracking the reinstatement of memory-related brain activity patterns during retrieval. The basic premise that content-specific neural patterns are reactivated during retrieval has been confirmed using fMRI (for reviews, see [9, 10]) and more recently also EEG and MEG [11-17]. However, no study has so far investigated the temporal fluctuations of memory-related patterns in human long-term memory, and whether they are systematically linked to brain oscillations.

A major computational challenge for our memory system is to effectively separate the information arriving from external sensory sources from the information generated in internal circuits. In other words, if the brain constantly pattern completes, how does it make sure that the neural coding of this internally (and possibly incorrectly) generated information does not interfere with the coding of new, incoming information? One promising explanation suggests that this is accomplished by means of neural oscillations. In particular, it has been argued that the phase of the hippocampal theta oscillation supports the chunking of mnemonic information such that the neural assemblies involved in encoding and retrieval are temporally segregrated [18, 19]. In a seminal paper, Pavlides, et al. [20] showed that stimulating a hippocampal assembly at one phase of the theta rhythm induced long-term potentiation

(LTP), whereas stimulating at the opposite phase induced long-term depression (LTD). This finding has since been replicated many times in rodents [21, 22], and implemented in computational models of episodic memory and the hippocampus [19, 23-26]. These models share the assumption that successful retrieval is most likely at one specific phase of the hippocampal theta rhythm, opposing the optimal encoding phase [19, 27]. Memory retrieval should be a continuously oscillating process that is locked to the hippocampal theta phase .

Direct evidence for theta phase modulation in human long-term memory still remains elusive. FMRI studies by nature are blind to the sub-second temporal dynamics that could mediate memory reinstatement, and electrophysiological studies have so far not investigated rhythmic fluctuations in memory reactivation. To our knowledge, only one previous study exists that has shown evidence for periodic reactivation, and this was during a working memory task [28]. In human long-term memory it is therefore unknown whether neural signatures of memory reactivation are locked to a theta rhythm. The present study was aimed at directly testing this hypothesis. EEG data was recorded while participants encoded novel word-object associations, and were later cued with the words to retrieve the objects. EEG-based pattern classifiers were trained to detect memory-related neural patterns during recall with high temporal precision. We demonstrate that within each retrieval period, classifier fidelity fluctuates at 7-8Hz within each retrieval period, and that this index of memory reactivation is locked to a particular phase of the same theta rhythm.

Results

Participants retrieve the episodic memories with high accuracy

The paradigm was a simple word-object associative memory task designed to yield a high number of correct trials (Figure 1A). Participants studied associations between action verbs and objects in random pairings, and were later cued with the word to retrieve the object. Two

measures of memory accuracy confirmed that participants performed the task well. The first was a subjective measure where participants indicated, via a button press after cue onset, whether and when they recalled the associated object. Participants on average indicated that they remembered the object on 94.21% (SD = 5.75%) of the trials. A second, more objective measure was accuracy in response to a question about the object's semantic category (animate vs inanimate), which appeared at the end of each retrieval trial, and which participants answered correctly on 88.20% (SD = 6.57%) of the trials. These two measures were highly correlated ($r_{Spearman}$ = 0.60, $p < .05$). Average accuracy for perceptual detail (photograph vs line drawing) was 85.31% (SD = 6.45%).

Reaction times for the first button press when retrieving animate (Mean = 3.03 secs, SD = .95 secs, min = 1.28 secs, max = 6.01 secs) and inanimate (Mean = 2.96 secs, SD = .77 secs, min = 1.47 secs, max = 4.24 secs) objects did not differ significantly, t(1,23) = .57, p = .58. The time window used for classification (-200ms to 1500ms around the cue) thus only minimally overlapped with the button press window.

Power spectrum of classifier shows strongest effects in lower frequencies

Our primary goal was to test whether the neural signatures of memory retrieval wax and wane in a theta oscillatory rhythm. Our neural index of memory retrieval was obtained from a linear discriminant analysis (LDA) trained to detect evidence for the reactivation of the correct object category (animate vs. inanimate) during retrieval (Figure 1B, see methods for details). The LDA was trained and tested independently per participant at each retrieval time point starting with the onset of the word cue, using a leave-one-out procedure. The input into the LDA was a feature vector containing the signal amplitudes from all 128 EEG channels at a given time point. The major output of interest was the fidelity (distance, or d-) values

available for each trial and time point. These values represent the distance from the hyperplane that optimally separates the two classes of retrieved objects (animate vs inanimate), and their timecourses served as our time-resolved, parametric index of memory reactivation. For the purpose of this study, the LDA was trained and tested during cued recall in order to isolate a purely retrieval-based signature or memory retrieval, which could then (below) be compared with a purely encoding-based index of memory classification. Additional analyses using classifiers trained on encoding and tested at retrieval are reported in the supplementary materials (Figure S1 and S4).

We first asked whether evidence could be found for an oscillation in these time-resolved indices of memory reactivation (Figure 2A-B). Fidelity timecourses from the recall task were averaged across trials per participant and subjected to a Fourier Transformation. If memory reactivation fluctuates in a theta rhythm, the resulting power spectra will show a selective increase in a band-limited lower (theta) frequency band. We compared the power spectra obtained from the real classifier outputs with a bootstrapped baseline [29], the latter using the d-value outputs from classifiers that were trained and tested on the same EEG trials but with randomly shuffled category labels (see Method section). This procedure controls for spurious power peaks that are driven by the frequency characteristics of the raw data (e.g. a dominant oscillation in the single trials). Significant power differences between the real and shuffled data were found in frequency bins at 7-9Hz and 13Hz, all exceeding the 95^{th} percentile of the empirical null distribution (Figure 2C). Power at 7-9Hz was significantly higher ($t(1,23) = 1.9425$, $p = .03$) when including only correctly retrieved trials that when including all trials, suggesting a relationship of the classifier fluctuation to memory success [30]. An alternative method with more stringent criteria to determine the presence of oscillations [31] confirmed that oscillatory power in the classifier time series was increased above baseline in the 7-9Hz frequency range (Figure 2D). Moreover, a similar power spectrum was found when the

classifier was trained on encoding and tested on retrieval (Figure S4.) The frequency characteristics of the classifier fidelity time courses thus suggest a rhythmic fluctuation in memory reactivation that was most consistent in the 7-9Hz frequency range.

Phase-amplitude coupling reveals oscillating patterns at retrieval for 8Hz

Our next two analyses were aimed at specifically testing for coupling between neural reactivation (i.e., classifier timeseries) and the phase of hippocampal theta-band oscillations. For this purpose, the raw EEG trials were projected into source space using an LCMV beamforming algorithm [32, 33], and a hippocampal mask was used to extract the 8Hz phase of the hippocampal virtual channels for each trial and time point. We computed a phase modulation index (MI) [35] reflecting the strength of coupling between the hippocampal 8Hz phase and the amplitude of the classifier output. Classifier fidelity as a function of hippocampal theta phase is plotted in Figure 2E (green line). This analysis revealed a significant modulation index ($M = .0071$, $SD = .0042$; baseline: $M = .0056$, $SD = .0006$), $t(1,23) = 1.8191$, $p < .05$, one-sided t-test, indicating that fidelity of the retrieval classifier was modulated by the phase of the hippocampal 8Hz oscillation (Figure 2E).

We next directly compared the theta phase at which classifier fidelity was maximal during encoding and retrieval. All basic analysis steps were repeated for the encoding EEG data, where an LDA discriminating animate from inanimate objects was trained and tested at each time point from 200ms before until 1500ms after object onset. The full time generalization matrices showing classifier performance for encoding and retrieval can be found in Figure S1. The 8Hz phase at encoding was then extracted from hippocampal virtual channels to calculate the phase modulation index. Classifier fidelity as a function of hippocampal theta phase during encoding is shown in Figure 2E (grey line). A significant phase modulation was found also for encoding ($M = .0068$, $SD = .0029$; baseline: $M = .0052$, $SD = .0007$), $t(1,23) =$

2.7494, $p < .05$, one-sided t-test). In order to directly compare the encoding and retrieval phases, we identified the phase at which encoding or retrieval classification was optimal in each subject. A Rayleigh circular statistic comparing the absolute phase angles at which encoding and retrieval classification was maximal revealed that these angles significantly differed from each other, $z(1,23) = 5.5342$, $p = .001$. Similar statistics were obtained by fitting a sine wave to the data and identifying and extracting the phase at which classification was optimal. Together, the results of the phase modulation analyses show that retrieval fluctuates as a function of hippocampal theta (8Hz) phase, and that the optimal retrieval phase is on average 188 degrees phase shifted compared with the optimal phase during encoding.

Classifier-locked averages reveal a consistent theta phase prior to memory reinstatement

Having established that the neural retrieval patterns oscillate and are coupled to an 8Hz oscillation, we next investigated the temporal relationship between theta phase and memory reinstatement. The analysis was inspired by the use of spike-triggered averages in animal intracranial work [34, 35]. We here adopted a similar approach computing *classifier-locked averages* around the time points of maximal memory reactivation (see Methods for details). On each single trial, those time points of maximal classifier fidelity that exceeded the 95^{th} percentile of a bootstrapped baseline were marked as new events of interest, the corresponding time stamps were located in the raw EEG epochs, and the ongoing EEG signal surrounding these maxima was then analysed for phase consistency across all electrodes (Figure 3A). We used a non-parametric cluster-based permutation test to compare the real data with a temporally shuffled baseline that keeps the EEG trial structure intact but produces a random temporal alignment between the classifier maxima and the ongoing phase (see Method section). Comparing the "real" times of maximum classifier fidelity with the

temporally shuffled baseline revealed a cluster of significant ($p_{corr} < .05$) phase consistency from 500ms to 50ms before the classifier maxima, centred at 7Hz (Figure 3B). Note that in this analysis, several classifier peaks per trial can exceed the 95th percentile criterium, and many of the classifier-locked EEG epochs will thus overlap, resulting in temporal smearing of the phase-locked activity. When running the same analysis extracting only one maximum per trial (Figure S2C), we found a similar cluster of phase locking but with a more narrow temporal extent from 500ms to 150ms pre-maxima, suggesting that the stongest phase-consistency effect was present roughly two theta cycles (corresponding to 2*143ms = 286ms) before mnemonic information could most confidetly be decoded. This finding supports our primary hypothesis that memory reinstatement shows a consistent oscillatory timing across trials and participants, in the same 7-9Hz frequency band at which the classifier fluctuates (Figure 2C).

It might seem counterintuitive that the strongest phase consistency was observed prior to the time points of maximum classification fidelity, rather than at the maxima themselves. However, this temporal relationship is to be expected if the phase-locked signal originates from a different, upstream region in the processing hierarchy compared to the signal that the classifier's decision is based on. Our findings are consistent with a model where the re-instantiation of a memory trace is triggered at a consistent phase of a hippocampal/MTL theta oscillation, followed by memory reinstatement in a broader range of neocortical regions representing the stored memory [1, 24, 36, 37]. The aim of the next analysis was to identify the brain regions involved in producing the observed clusters of theta phase consistency, with the hypothesis that the effect should be present in MTL areas [1, 2].

Trial time-courses were projected into source space using a beamforming algorithm [33, 38], and we then looked for the sources showing the strongest phase consistency. Contrasting all classifier maxima with the shuffled baseline (identical to the scalp level analysis), we found

an activation cluster spanning large regions of occipital, temporal and frontal cortex, primarily in the right hemisphere (maximum at MNI coordinates xyz = 10 -10 10, Thalamus, Figure 3C). While these sources included medial temporal lobe areas, they do not suggest a specific role of the hippocampus in producing the theta phase-locked signal preceding the classifier maxima.

High classifier fidelity is associated with strong theta phase consistency in MTL

We next wanted to test whether the theta phase consistency systematically varied with the strength of neural reinstatement. We hypothesized that phase consistency would be highest when the classifier correctly detected neural reactivation with high fidelity, and lower when the classifier was correct, but less confident.

Comparing classifier maxima of higher and lower fidelity revealed a significant ($p_{corr} < .05$) cluster at 7Hz preceding the maxima by 500ms to 200ms (Figure 3D). This cluster highly overlapped in timing, frequency and topography with our previous classifier-triggered average analyses. When conducting the same analysis in source space, we found sources that spanned the parietal and the right medial temporal lobes (maximum MNI coordinates xyz = 50 -30 30, inferior parietal lobule; Figure 3E), strongly reminiscent of the core recollection or memory success network typically found in fMRI studies [39]. Our data thus suggest that the neural signatures of memory retrieval are linked to a specific phase of a theta oscillation, and this phase relationship becomes stronger with more confident neural reactivation. The source level analysis additionally confirms our a priori assumption that the phase-locked signal that precedes memory reactivation involves the MTL and other core recollection areas.

Theta phase-locking is unlikely to be produced by early cue-related effects

While consistent with our hypotheses, this pattern of results could in theory also be explained by an early ERP elicited by the reminder word, since ERPs are generally associated with

strong phase locking in slow frequencies [40]. Such an explanation would assume that our classifier maxima tend to occur at a consistent time point within each retrieval trial with a delay to the reminder-elicited ERP of approximately 300ms. Several observations speak against this alternative. First, the classifier maxima were relatively evenly distributed across the entire retrieval period and did not tend to cluster around early time points. A slight increase in density was found in the typical recollection time window [41] from 400-800ms post-cue, but the overall distribution of the maxima did not significantly differ from uniform (χ^2 = (1, N = 6007) = 7376600, p = .375) (Figure S2A). Second, we repeated the classifier-triggered average analysis excluding all classifier maxima that occurred earlier than 400ms or 600ms post-cue, respectively, excluding the time delays that would be most strongly affected by early ERPs. Both analyses revealed a significant phase-locking effect ($p_{corr} < .05$) in a very similar time window and frequency band as in the original analysis (Figure S2E and F). This result indicates that the theta phase-locked process preceding memory reinstatement can occur at various delays in a recall trial.

EEG signals at the exact time points of maximal classifier fidelity show content-dependent differences with a source in anterior temporal lobe

In order to correctly classify a trial as belonging to one category or another, linear classifiers including LDA require a consistent EEG signal difference across trials. If these signal differences additionally have consistent timing and topography across participants, we should on average be able to observe a robust signal difference between animate and inanimate objects at time points of confident classification. We therefore conducted two confirmatory ERP analyses comparing the average waveforms for animate and inanimate objects during retrieval. The first of these analyses contrasted animate and inanimate trials time-locked to the onset of the word cue (Figure S3A-B). This analysis shows that the strongest average signal differences were present over frontal channels, although this cluster did not survive

correction for multiple comparisons using cluster-based permutation statistics (p_{corr} = .64). The lack of significance could be due to variance in retrieval latency across trials, varying topographies across participants, or in fact due to an oscillating process that makes it difficult to observe a coherent cluster in time. Interestingly, when conducting an FFT on the average ERP differentiating animate and inanimate object retrievals in each participant, these signal differences showed power increases above baseline at 6-9Hz (Figure S3C), in the same range revealed by our frequency transformation of the classifier fidelity values. This finding confirms that the 8Hz oscillation is inherent in the signal difference that the LDA relies on.

The second ERP analysis again contrasted animate and inanimate trials, but this time locked to the time points of maximal classifier fidelity (as used in previous analyses). A cluster-based permutation test revealed a significant cluster ($p_{corr} < .05$) over frontal electrodes, spanning from 90ms before to 120ms after the classifier maxima (Figure 4A). Reconstructed at source level (Figure 4B), this effect showed a maximum in left anterior temporal lobe (maximum MNI coordinates xyz = -30 10 -40, superior temporal gyrus; and -40 0 -40, inferior temporal gyrus). The results confirm that the single-trial classifier maxima indeed reflect a meaningful difference in the neural patterns elicited by retrieving different types of objects, rather than reflecting random fluctuations in classifier performance. The most likely source of the effect was found in anterior temporal lobe, an area strongly linked to semantic memory processing [42, 43], where previous studies found tight links between classifier fidelity and the speed at which participants behaviourally categorize objects as animate or inanimate [38]. Together, the two ERP analyses validate our LDA approach and provide converging evidence that retrieval-related differences between animate and inanimate objects fluctuate in the theta range and are most pronounced over neocortical regions involved in high-level semantic processing [44].

Classifiers that generalise from encoding to retrieval show similar frequency characteristics

The results reported so far focus on an index of memory reactivation derived from classifiers trained and tested on the retrieval data. Below, we report additional analyses conducted on classifiers that were trained on the encoding data, and then tested either on the encoding or on the retrieval data. Encoding-to-retrieval classification has been commonly used in previous studies [12]. We conducted the additional analyses to confirm that such classifiers can also successfully detect memory reactivation, and that their frequency characteristics are similar to our main, purely retrieval-based metric. The results are summarized in Figure S4.

Encoding analyses were conducted on epochs time-locked to the onset of the animate and inanimate objects (-200ms to 1500ms). As a first step, an LDA was trained on encoding and also tested during encoding (Figure S4A). In line with the existing literature on object perception [58], animate vs inanimate category membership could be best decoded in a time window around 300ms after object onset, with an accuracy peak at 305ms. The classifier fidelity timecourses were then averaged within participants and subjected to a Fourier Transformation, following the same procedure as for the retrieval data. The resulting spectra (Figure S4B) showed the strongest power in lower frequencies with peaks at 3, 5, and 6 Hz exceeding the 95^{th} percentile of the random label chance distribution.

During the time window where the LDA performed best, we also found a univariate ERP cluster ($p_{corr} < .05$) from 240-340ms with a frontal topography that significantly differentiated animate from inanimate objects during encoding (Figure S4C). Note that this cluster had a frontal topography similar to the main cluster differentiating animate from inanimate objects during retrieval (as shown in Figure S3A), providing a first indication that content-specific processes engaged during encoding might be re-engaged during retrieval.

Based on this observation, we next tested explicitly whether classifiers trained to distinguish animate from inanimate objects during encoding could successfully discriminate those categories during retrieval. For this analysis, the classifier was trained on each time point within the 240-340ms encoding interval identified above, and tested at each time point at retrieval (see Figure S4D). This approach revealed the highest decoding accuracy in a retrieval time window from approximately 800-1500ms, a window typically associated with successful recollection [41]. We then assessed the frequency characteristics of the encoding-retrieval classifiers using the same FFT method as before, but this time applied to the classifiers trained on the activity patterns between 240-340ms during encoding, and tested at each time point during retrieval (see Figure S4E). The resulting power spectra showed the maximum peak at 9Hz (5 and 9Hz exceeding the 95th percentile), with a similar distribution but at a slightly higher frequency peak compared with results obtained when training and testing at retrieval (see main Figure 2C).

Discussion

Memory retrieval, or at least the neural reactivation process underlying it, is often thought of as a static process that happens in an all-or-none fashion once a reminder has reactivated a past experience. However, evidence from rodents suggests that pattern completion fluctuates on a sub-second time scale, and that these fluctuations are determined by a hippocampal theta oscillation that shifts the network between states optimal for encoding, and states optimal for retrieval [19, 23]. We here sought to investigate these oscillating retrieval dynamics in humans in a cued recall task. Several findings from the present experiment indicate that the neural signatures of memory reactivation in fact do fluctuate within a single recall trial in the human brain, and are tightly linked to a specific phase of a theta oscillation.

Our main metric of interest was a parametric, time-resolved index of memory reactivation for each trial that we obtained from a multivariate classifier trained to detect the semantic category of the recalled object. First, we found that this index in itself fluctuates at 7-8Hz. This oscillating pattern was evident in the average classifier fidelity time courses from each participant (Figure 2C), relative to a baseline which used the output from random label classifiers. The effect can thus not readily be explained by the frequency structure of the data that served as input to the classifier (e.g., a dominant 7-8Hz rhythm inherent in the EEG epochs). The 7-9Hz fluctuation was stronger for successfully remembered than for all associations including misses, and it was also present in the average ERP waveforms differentiating the retrieval of animate and inanimate objects (Figure S3C). These findings suggest a fluctuation in the signals differentiating the two classes of retrieved mnemonic representations, consistent with a rhythmic memory reactivation process.

Second, we investigated whether the classifier-based indices of memory reactivation systematically varied as a function of theta phase. We found that classifier fidelity was significantly modulated by the phase of an 8Hz oscillation extracted from virtual hippocampal channels (Figure 2E). The phase of peak classification fidelity during recall was 188 degrees shifted compared to the phase of peak fidelity during encoding. These results support two of the central claims of the Hasselmo model: that neural signatures of memory reactivation are tighly coupled to a particular phase of a hippocampal 8Hz oscillation; and that the optimal phase for memory retrieval is flipped relative to the optimal encoding phase along this same theta oscillation [19].

Third, to scrutinize the temporal relationship between memory retrieval and theta phase, we tested whether the time points where our classifier indicated maximal neural memory reinstatement were time-locked to a consistent phase in the same frequency range, as would be the case if retrieval was initiated at a particular theta phase. A classifier-locked EEG

analysis, inspired by animal work, revealed significant phase alignment at 7-8Hz, preceding the time points of maximal memory reactivation by approximately 200-300ms (Figure 3B-C). This cluster remained robust irrespective of whether we included only one classifier maximum or several maxima per trial (Figure S2C), when including correct trials only (Figure S2D), and when excluding early maxima close to the onset of the word cue (Figure S2E-F). Together, these findings suggest a close functional relationship between the phase of an ongoing theta oscillation, and neural memory reinstatement as measured by EEG classifiers, in line with the computational models that motivated our hypotheses [19, 23, 45].

The functional coupling between memory reinstatement and oscillatory phase is further corroborated by an analysis that contrasted phase consistency between classifier maxima of high and low fidelity, used as a proxy for strong vs weak memory reactivation (Figure 3D-E). Phase consistency in the 7-8Hz frequency and -500 to -200ms time range was higher for high-fidelity trials. The sources producing the difference between high and low fidelity maxima spanned medial and lateral parietal regions, and medial temporal lobe areas including the hippocampus. These regions are typically engaged during successful recollection [39] and show strong functional connectivity with the hippocampus [46]. While we cannot establish the hippocampus as a unique source of the theta phase-locking effect, our results are at minimum consistent with a hippocampal theta oscillation that extends into the functionally connected core recollection network. A link to medial temporal is also corroborated by the first analysis showing modulation of memory reactivation by the hippocampal 8Hz phase (Figure 2E). Together with the phase-locking results, our findings thus support theories suggesting that episodic memory retrieval relies on periodic cycles of communication between storage/retrieval systems in medial temporal lobe and neocortical areas that represent the various components of an episode [1, 2].

The exact time course of the interaction between hippocampus and neocortex during retrieval is still not fully understood. Electrophysiological studies using time-resolved multivariate methods have detected memory reactivation in the typical recollection time window [11, 16, 17]. Consistent with this timing, our classifier maxima had a tendency to cluster in the recollection window around 400-800ms post-cue (Figure S2A). Our main interest in this study, however, was whether neural reactivation was linked to a consistent oscillatory phase in the theta band irrespective of when exactly it is triggered within a trial. Our findings provide strong evidence for such phasic modulation within a recall trial, in line with models suggesting that memory retrieval is initiated at an optimal phase of a hippocampal theta oscillation [19].

At the exact time of the classifier maxima, we observed a significant difference in the ERPs distinguishing between the different types of retrieved memories (i.e., animate vs inanimate, Figure 4). The main source of this difference was localized to the anterior temporal lobe, consistent with this region's role in representing abstract object information [42]. Note that it is not surprising that we observed such an ERP effect, since the classifier requires a reliable signal difference in order to detect differences in reactivated content. The source of this signal is interesting, however, indicating that the classifier's decisions are based on information that originates from neocortical sources that are likely to represent the reactivated memory's content, and have little overlap with the sources of the theta phase-locked signal. Overall, our findings suggest that a few hundred milliseconds before the brain reinstates a memory in neocortex, an oscillating process in the MTL initiates retrieval, leading to a memory signal that oscillates and is modulated by the hippocampal theta phase [2, 5].

To our knowledge, our study is the first that directly links memory reinstatement to theta phase in human long-term memory. Previous studies have investigated the role of theta phase in working memory, and have provided first evidence for a phase shift between encoding and

retrieval [47]. They also suggest that theta phase plays a role in orchestrating gamma (30-80Hz) oscillations during periods of working memory maintenance [48, 49]. High frequency activity in the gamma range is thought to represent the firing of cell assemblies that code for the content of mental representations, and lower frequencies presumably provide the time windows for the firing of these assemblies [18, 28, 48-50]. Following this logic, Fuentemilla, et al. [28] used a delayed match-to-sample working memory task to investigate how gamma patterns representing the encoded material re-emerged during maintenance. Reactivation took place several times over a 5-sec delay, and these reactivations were phase-locked to a theta oscillation. Rodent work also suggests a link between gamma oscillations and theta phase. Different hippocampal subfields produce faster or slower gamma oscillations depending on whether the animal is encoding novel information or retrieving familiar information, and these two gamma rhythms are coupled to distinct phases of the hippocampal theta rhythm [51]. Our results provide the first evidence for a similar relationship in human long-term memory, using a classifier-based metric rather than gamma oscillations as a proxy for memory reinstatement and its relationship to the ongoing EEG.

We hope that our method will prove useful as a general approach for probing the relationship between information coding and the phase of slow oscilaltions. Phase coding has been suggested as an important mechanism outside the memory domain, including attentional selection [52] and spatial navigation [53]. Within memory, our approach could be used to directly test whether distinct parts of a sequence of events are represented at different phases along a theta oscillation [54], or whether memories are reactivated at specific phases of slow oscillations during sleep [55, 56]. Computational models [57] also postulate that phase coding is crucial for resolving mnemonic competition when several memories are simultaneously reactivated by a reminder. Building on our method and findings, follow-up studies can directly test phase coding as a mechanism of organizing memories (e.g. according to their

relevance) during encoding, during offline periods following encoding, and when reactivating memories during retrieval.

In sum, the present experiment shows that memories – or their neural signatures – wax and wane on a millisecond time scale within a trial, and that their neural reactivation follows the phase of a 7-8Hz theta rhythm. These findings provide the first direct support for theta phase encoding-retrieval models in the human brain, and thus bridge an important gap between computational, rodent and human work.

Acknowledgements

This work was supported by a fellowship from Stiftelsen Olle Engkvist Byggmästare awarded to M.W. and C.K., and a Starting Grant from the European Research Council awarded to M.W. (ERC-2016-STG-715714). We also thank James Lloyd-Cox for help with data acquisition, and Matthias Treder, Benjamin Griffiths and Sebastian Michelmann for their helpful conceptual input during data analysis.

Author Contribution

Conceptualization, C.K., J.L.-D. and M.W.; Methodology, C.K., J.L.-D., S.H., and M.W.; Investigation J.L.-D.; Formal Analysis, C.K. and J.L.-D.; Writing – Original Draft, C.K. and M.W.; Writing – Review & Editing, C.K., J.L.-D., S.H., and M.W.; Visualization, C.K.; Funding acquisition, C.K. and M.W.

Declaration of Interests

The authors declare no competing financial interests.

References

1. McClelland, J.L., McNaughton, B.L., and O'Reilly, R.C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychol Rev *102*, 419-457.
2. Teyler, T.J., and DiScenna, P. (1986). The hippocampal memory indexing theory. Behav Neurosci *100*, 147-154.
3. Tulving, E., and Markowitsch, H.J. (1998). Episodic and declarative memory: role of the hippocampus. Hippocampus *8*, 198-204.
4. Rolls, E.T. (1996). A theory of hippocampal function in memory. Hippocampus *6*, 601-620.
5. Teyler, T.J., and Rudy, J.W. (2007). The hippocampal indexing theory and episodic memory: updating the index. Hippocampus *17*, 1158-1169.
6. Marr, D. (1971). Simple memory: a theory for archicortex. Philos Trans R Soc Lond B Biol Sci *262*, 23-81.
7. Alvarez, P., and Squire, L.R. (1994). Memory consolidation and the medial temporal lobe: a simple network model. Proc Natl Acad Sci U S A *91*, 7041-7045.
8. Norman, K.A., and O'Reilly, R.C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. Psychol Rev *110*, 611-646.
9. Danker, J.F., and Anderson, J.R. (2010). The ghosts of brain states past: remembering reactivates the brain regions engaged during encoding. Psychol Bull *136*, 87-102.
10. Rissman, J., and Wagner, A.D. (2012). Distributed representations in memory: insights from functional brain imaging. Annu Rev Psychol *63*, 101-128.
11. Johnson, J.D., Price, M.H., and Leiker, E.K. (2015). Episodic retrieval involves early and sustained effects of reactivating information from encoding. Neuroimage *106*, 300-310.
12. Waldhauser, G.T., Braun, V., and Hanslmayr, S. (2016). Episodic Memory Retrieval Functionally Relies on Very Rapid Reactivation of Sensory Information. J Neurosci *36*, 251-260.
13. Wimber, M., Maass, A., Staudigl, T., Richardson-Klavehn, A., and Hanslmayr, S. (2012). Rapid memory reactivation revealed by oscillatory entrainment. Curr Biol *22*, 1482-1486.
14. Kurth-Nelson, Z., Barnes, G., Sejdinovic, D., Dolan, R., and Dayan, P. (2015). Temporal structure in associative retrieval. Elife *4*.
15. Staudigl, T., Vollmar, C., Noachtar, S., and Hanslmayr, S. (2015). Temporal-pattern similarity analysis reveals the beneficial and detrimental effects of context reinstatement on human memory. J Neurosci *35*, 5373-5384.
16. Jafarpour, A., Fuentemilla, L., Horner, A.J., Penny, W., and Duzel, E. (2014). Replay of Very Early Encoding Representations during Recollection. The Journal of Neuroscience *34*, 242-248.
17. Michelmann, S., Bowman, H., and Hanslmayr, S. (2016). The Temporal Signature of Memories: Identification of a General Mechanism for Dynamic Memory Replay in Humans. PLoS Biol *14*, e1002528.
18. Nyhus, E., and Curran, T. (2010). Functional role of gamma and theta oscillations in episodic memory. Neurosci Biobehav Rev *34*, 1023-1035.
19. Hasselmo, M.E., Bodelon, C., and Wyble, B.P. (2002). A proposed function for hippocampal theta rhythm: separate phases of encoding and retrieval enhance reversal of prior learning. Neural Comput *14*, 793-817.
20. Pavlides, C., Greenstein, Y.J., Grudman, M., and Winson, J. (1988). Long-term potentiation in the dentate gyrus is induced preferentially on the positive phase of theta-rhythm. Brain Res *439*, 383-387.
21. Huerta, P.T., and Lisman, J.E. (1993). Heightened synaptic plasticity of hippocampal CA1 neurons during a cholinergically induced rhythmic state. Nature *364*, 723-725.
22. Hyman, J.M., Wyble, B.P., Goyal, V., Rossi, C.A., and Hasselmo, M.E. (2003). Stimulation in hippocampal region CA1 in behaving rats yields long-term potentiation when delivered to

the peak of theta and long-term depression when delivered to the trough. J Neurosci *23*, 11725-11731.
23. Kunec, S., Hasselmo, M.E., and Kopell, N. (2005). Encoding and retrieval in the CA3 region of the hippocampus: a model of theta-phase separation. J Neurophysiol *94*, 70-82.
24. Buzsaki, G. (2002). Theta oscillations in the hippocampus. Neuron *33*, 325-340.
25. Hasselmo, M.E., and Eichenbaum, H. (2005). Hippocampal mechanisms for the context-dependent retrieval of episodes. Neural Netw *18*, 1172-1190.
26. Parish, G., Hanslmayr, S., and Bowman, H. (2017). The Sync/deSync model: How a synchronized hippocampus and a de-synchronized neocortex code memories. bioRxiv.
27. Hasselmo, M.E. (2005). What is the function of hippocampal theta rhythm?--Linking behavioral data to phasic properties of field potential and unit recording data. Hippocampus *15*, 936-949.
28. Fuentemilla, L., Penny, W.D., Cashdollar, N., Bunzeck, N., and Duzel, E. (2010). Theta-coupled periodic replay in working memory. Curr Biol *20*, 606-612.
29. Stelzer, J., Chen, Y., and Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): random permutations and cluster size control. Neuroimage *65*, 69-82.
30. Siegel, M., Warden, M.R., and Miller, E.K. (2009). Phase-dependent neuronal coding of objects in short-term memory. Proc Natl Acad Sci U S A *106*, 21341-21346.
31. Watrous, A.J., Miller, J., Qasim, S.E., Fried, I., and Jacobs, J. (2017). Phase-tuned neuronal firing encodes human contextual representations for navigational goals. bioRxiv.
32. Van Veen, B.D., van Drongelen, W., Yuchtman, M., and Suzuki, A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. IEEE Trans Biomed Eng *44*, 867-880.
33. Gross, J., Kujala, J., Hamalainen, M., Timmermann, L., Schnitzler, A., and Salmelin, R. (2001). Dynamic imaging of coherent sources: Studying neural interactions in the human brain. Proc Natl Acad Sci U S A *98*, 694-699.
34. Douchamps, V., Jeewajee, A., Blundell, P., Burgess, N., and Lever, C. (2013). Evidence for encoding versus retrieval scheduling in the hippocampus by theta phase and acetylcholine. J Neurosci *33*, 8689-8704.
35. Rieke, F., Warland, D., de Ruyter van Steveninck, R., and Bialek, W. (1997). Spikes: Exploring the neural code, (Cambridge, MA).
36. O'Reilly, R.C., and Norman, K.A. (2002). Hippocampal and neocortical contributions to memory: advances in the complementary learning systems framework. Trends Cogn Sci *6*, 505-510.
37. O'Reilly, R.C., Bhattacharyya, R., Howard, M.D., and Ketz, N. (2014). Complementary learning systems. Cognitive science *38*, 1229-1248.
38. Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci *2011*, 156869.
39. Rugg, M.D., and Vilberg, K.L. (2013). Brain networks underlying episodic memory retrieval. Curr Opin Neurobiol *23*, 255-260.
40. Gruber, W.R., Klimesch, W., Sauseng, P., and Doppelmayr, M. (2005). Alpha phase synchronization predicts P1 and N1 latency and amplitude size. Cereb Cortex *15*, 371-377.
41. Yonelinas, A.P. (2002). The Nature of Recollection and Familiarity: A Review of 30 Years of Research. Journal of Memory and Language *46*, 441-517.
42. Patterson, K., Nestor, P.J., and Rogers, T.T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. Nat Rev Neurosci *8*, 976-987.
43. Duvernoy, H., Cattin, F., and Risold, P. (2007). The Human Hippocampus. Functional Anatomy, Vascularization and Serial Sections with MRI., 4th Edition, (Springer).

44. Tyler, L.K., Chiu, S., Zhuang, J., Randall, B., Devereux, B.J., Wright, P., Clarke, A., and Taylor, K.I. (2013). Objects and categories: feature statistics and object processing in the ventral stream. J Cogn Neurosci *25*, 1723-1735.
45. Ketz, N., Morkonda, S.G., and O'Reilly, R.C. (2013). Theta Coordinated Error-Driven Learning in the Hippocampus. PLOS Computational Biology *9*, e1003067.
46. Wang, J.X., Rogers, L.M., Gross, E.Z., Ryals, A.J., Dokucu, M.E., Brandstatt, K.L., Hermiller, M.S., and Voss, J.L. (2014). Targeted enhancement of cortical-hippocampal brain networks and associative memory. Science *345*, 1054-1057.
47. Rizzuto, D.S., Madsen, J.R., Bromfield, E.B., Schulze-Bonhage, A., and Kahana, M.J. (2006). Human neocortical oscillations exhibit theta phase differences between encoding and retrieval. Neuroimage *31*, 1352-1358.
48. Jensen, O. (2006). Maintenance of multiple working memory items by temporal segmentation. Neuroscience *139*, 237-249.
49. Fell, J., and Axmacher, N. (2011). The role of phase synchronization in memory processes. Nat Rev Neurosci *12*, 105-118.
50. Tallon-Baudry, C., and Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. Trends Cogn Sci *3*, 151-162.
51. Colgin, L.L., Denninger, T., Fyhn, M., Hafting, T., Bonnevie, T., Jensen, O., Moser, M.B., and Moser, E.I. (2009). Frequency of gamma oscillations routes flow of information in the hippocampus. Nature *462*, 353-357.
52. Jensen, O., Bonnefond, M., and VanRullen, R. (2012). An oscillatory mechanism for prioritizing salient unattended stimuli. Trends Cogn Sci *16*, 200-206.
53. O'Keefe, J., and Recce, M.L. (1993). Phase relationship between hippocampal place units and the EEG theta rhythm. Hippocampus *3*, 317-330.
54. Heusser, A.C., Poeppel, D., Ezzyat, Y., and Davachi, L. (2016). Episodic sequence memory is supported by a theta-gamma phase code. Nat Neurosci *19*, 1374-1380.
55. Hanert, A., Weber, F.D., Pedersen, A., Born, J., and Bartsch, T. (2017). Sleep in Humans Stabilizes Pattern Separation Performance. J Neurosci *37*, 12238-12246.
56. Staresina, B.P., Bergmann, T.O., Bonnefond, M., van der Meij, R., Jensen, O., Deuker, L., Elger, C.E., Axmacher, N., and Fell, J. (2015). Hierarchical nesting of slow oscillations, spindles and ripples in the human hippocampus during sleep. Nat Neurosci *18*, 1679-1686.
57. Norman, K.A., Newman, E., Detre, G., and Polyn, S. (2006). How inhibitory oscillations can train neural networks and punish competitors. Neural Comput *18*, 1577-1610.
58. Brodeur, M.B., Dionne-Dostie, E., Montreuil, T., and Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. PLoS One *5*, e10773.
59. Carlson, T., Tovar, D.A., Alink, A., and Kriegeskorte, N. (2013). Representational dynamics of object vision: the first 1000 ms. J Vis *13*.
60. Brainard, D.H. (1997). The Psychophysics Toolbox. Spat Vis *10*, 433-436.
61. Blankertz, B., Lemm, S., Treder, M., Haufe, S., and Muller, K.R. (2011). Single-trial analysis and classification of ERP components--a tutorial. Neuroimage *56*, 814-825.
62. Carlson, T.A., Ritchie, J.B., Kriegeskorte, N., Durvasula, S., and Ma, J. (2014). Reaction time for object categorization is predicted by representational distance. J Cogn Neurosci *26*, 132-142.
63. Voytek, B., Kramer, M.A., Case, J., Lepage, K.Q., Tempesta, Z.R., Knight, R.T., and Gazzaley, A. (2015). Age-Related Changes in 1/f Neural Electrophysiological Noise. J Neurosci *35*, 13257-13265.
64. Miller, K.J., Sorensen, L.B., Ojemann, J.G., and den Nijs, M. (2009). Power-law scaling in the brain surface electric potential. PLoS Comput Biol *5*, e1000609.

65. Tort, A.B., Komorowski, R., Eichenbaum, H., and Kopell, N. (2010). Measuring phase-amplitude coupling between neuronal oscillations of different frequencies. J Neurophysiol *104*, 1195-1210.
66. Cohen, M.X. (2014). Analyzing Neural Times Series Data: Theory and Practice., (MIT Press).
67. Lancaster, J.L., Woldorff, M.G., Parsons, L.M., Liotti, M., Freitas, C.S., Rainey, L., Kochunov, P.V., Nickerson, D., Mikiten, S.A., and Fox, P.T. (2000). Automated Talairach atlas labels for functional brain mapping. Hum Brain Mapp *10*, 120-131.
68. King, J.R., and Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. Trends Cogn Sci *18*, 203-210.
69. Cohen, M.X. (2014). Fluctuations in oscillation frequency control spike timing and coordinate neural networks. J Neurosci *34*, 8988-8998.
70. Lega, B.C., Jacobs, J., and Kahana, M. (2012). Human hippocampal theta oscillations and the formation of episodic memories. Hippocampus *22*, 748-761.

Main-text figure/table legends

Figure 1. Trial structure and Multivariate Pattern Analysis. (**A)** At encoding, participants associated action verbs with images depicting either an animate or inanimate object. After a short distractor task, participants were tested on the previously learned associations. The action verb was shown as a cue, asking participants to retrieve the associated object, and to indicate with a button press when the object came back to mind. They then had to respond to the question whether it was an animate or inanimate object. (**B)** For each time point and each trial from cue onset at retrieval, a linear discriminant analysis (LDA) classifier was trained and tested on detecting evidence for retrieval of the correct object category. The output of the classifier was a parametric value for each time point, reflecting the fidelity of the classifier to differentiate between the two object classes.

Figure 2. Analysis rationale and results of the time-frequency analyses relating classifier fidelity to theta oscillations and phase-modulation. (A) Example of a single-trial output from the LDA, reflecting the fidelity of the classifier in detecting the retrieved object's correct category at each time point during a retrieval trial. The black line represents a theta oscillation to illustrate our assumption that neural indices of memory reinstatement (i.e., the d-value time series) rhythmically fluctuate, and that the time points of maximal classifier fidelity should be consistently related to a particular phase of the underlying oscillation. **(B)** D-values were subjected to a Fourier transformation which reveals the power in each frequency band. **(C)** The resulting power spectrum shows significant deviations from an empirical null distribution at 7 to 9Hz and 13Hz. The baseline power spectrum was obtained from a combination of random label classifiers and bootstrapping, and is shown in grey (mean and SD). Values of the real classifier outputs exceeding the 95th percentile of the baseline distribution are marked as significant. **(D)** Frequency decomposition of the classifier time series using an alternative approach to detect frequencies [38], again showing above baseline power at slow frequencies including 7-8Hz. Figure showing mean ± SEM for baseline (grey lines) and 95th percentile (thick grey line). **(E)** Phase-amplitude coupling between EEG phase and classifier fidelity at source level revealed a significant modulation index averaged over hippocampal virtual channels (mask shown on the left) for 8Hz. Figure showing mean ± SD. See also Figure S4.

Figure 3. Rationale for classifier-locked average analysis to test for a functional relationship between classification maxima and neural memory reinstatement. **(A)** Classifier d-values exceeding the 95th percentile of the chance distribution were marked, corresponding time stamps were found in the ongoing EEG data, and the EEG was then re-epoched relative to the classifier maxima. This procedure resulted in new epochs with the classifier maxima at time zero. **(B)** Results of the classifier-locked average analysis relating classifier maxima to ongoing EEG phase. A non-parametric cluster-based permutation test revealed a significant cluster of phase consistency centred at 7-8 Hz, spanning from 500ms to 50ms before the maxima. **(C)** At source level, the maximal phase consistency was observed in occipital and right temporal lobe. (**D**) Contrasting maxima of high fidelity and maxima of lower fidelity, a significant cluster was again found at 7-8 Hz, from 500ms to 200ms before the maxima. **(E)** At source level, the maximal phase consistency effect was located in parietal and temporal lobes, including MTL, when contrasting high and low fidelity trials. Time-frequency plots highlight the significant cluster in time and frequency. Topographical and source level plots show values above the critical t-threshold (t-value of 1.7, 23 degrees of freedom, one-sided test) for significance. See also Figure S2.

Figure 4. Event-related potentials centred around classifier maxima, on scalp and source level. (A) ERPs locked to the time points of maximum classifier fidelity. A non-parametric cluster-based permutation test revealed a significant ($p < .05$,

cluster-corrected) difference in the average signal produced by animate and inanimate recall trials, confirming that a robust difference between retrieved object classes was present at the time points of maximum classifier fidelity. The ERP plot shows the average of the significant channels for descriptive purposes. **(B)** The classifier-locked ERP reconstructed at source level shows a maximum in anterior temporal lobe, regions assumed to be involved in high-level semantic processing. Source level plot show values above the critical t-threshold (t-value of 1.7, 23 degrees of freedom, one-sided test). See also Figure S3.

STAR Methods

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Casper Kerrén (C.Kerren@pgr.bham.ac.uk).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All experimental procedures in the present study were approved by and conducted in accordance with the University of Birmingham Research Ethics Committee (STEM). Written informed consent was obtained from participants before they took part in the experiment.

Participants

Twenty-four healthy participants (19 female) aged 18-32 years (mean = 22.1, SD = 4.7 years) received credits or monetary payment for participation. Participants had normal or corrected-to-normal vision and reported no history of neurological disorders.

METHOD DETAILS

Experimental Methods

Material and Setup

The material consisted of 64 images depicting animate objects (equal number of mammals, birds, insects, and marine animals) and 64 images depicting inanimate objects (equal number of electronic devices, clothes, fruits, and vegetables), taken from BOSS database [58] and

from online royalty-free databases, and was used due to previous success at distinguishing these categories using multi-variate pattern analysis [59]. All images were scaled to 500 x 500 pixels. A black-and-white drawing version of each image was manually created using GNU imaging manipulation software (www.gimp.org). The photographs vs. drawings served as an additional perceptual category (not of interest for the purpose of our current analyses). In addition to the material used for the experiment, 16 images were used for demonstrative purpose. Images from both semantic classes were randomly split into 16 sets, so that each set consisted of 8 images, 4 animate and 4 inanimate. Each set constituted one learning block. In addition, a list of 128 action verbs was generated for the experiment, serving as cue words in the cued recall task.

The experiment was set up via custom written MATLAB 2016a (©The Mathworks, Munich, Germany) code using functions from the Psychophysics Toolbox Version 3 [60]. The presentation was done on a 15-inch computer screen with Windows 64 bit.

Paradigm

Participants received instructions about the task and first performed two practice blocks. All participants then performed 16 experimental blocks (8 trials per block), each consisting of an associative learning phase, a distractor task, and a retrieval test (Figure 1). A learning trial consisted of a jittered fixation cross (between 500 and 1500ms), a unique action verb (1500ms), a fixation cross (between 500 and 1500ms), followed by a picture of an object that was presented in the centre of the screen for a minimum of 2 and a maximum of 10 seconds. The task was to come up with a vivid mental image that involved the object and the action verb presented in the current trial. As soon as they had a clear association in mind, participants pressed the up-arrow key on the keyboard, which led to the onset of the next trial.

Participants were aware of the later memory test, and knew that they had to pay attention to perceptual and meaningful aspects to perform the memory test.

A distractor task followed each learning phase. Here participants had to respond if a given random number (between 1 and 99) presented on the screen was odd or even. They were instructed to accomplish as many trials as they could in 45 seconds, and received feedback about their accuracy at the end of each distractor block.

After the distractor task, participants' memory for the 8 verb-object associations learned in the immediately preceding learning phase was tested in random order. Each trial consisted of a jittered fixation cross (500-1500ms), followed by one of the action verbs as a reminder cue for the association. Participants were asked to bring back to mind the object that had been associated with this word as vividly as possible. To capture the particular moment when participants consciously recalled a specific object, they were asked to press the up-arrow key as soon as they had a complete image of the associated memory in mind; or the down-arrow if they were unable to remember the association. The reminder was presented on the screen for a minimum of 2 seconds and until a response was made. Immediately following the button press, a blank square with the same size as the original images was displayed, and participants were asked to hold the retrieved object in mind for 3000ms. After a short fixation interval (1500ms), two questions were displayed sequentially, asking participants whether the associated object was a photograph or line-drawing (perceptual question), or an animate or inanimate object (semantic question). The order of questions was pseudo-random across trials such that the semantic question was asked first on half of the trials, and second on the other half.

Each semantic category was presented equally often in each type of perceptual level per participant. The action verbs were randomly assigned to the word-object pairs, and the

distribution of object categories for perceptual and semantic features was equally distributed across the first and second half of the experiment.

EEG Data Analysis

The electroencephalogram (EEG) was recorded using a BioSemi Active-Two Recording System (BioSemi, Amsterdam, the Netherlands) with a 128-channel electrode cap, sampled at 1024 Hz.

Preprocessing

Preprocessing was done twice using the FieldTrip toolbox [38] and custom written MATLAB code: First before implementing multivariate pattern analysis, and again after re-epoching the data based on the maxima of the classifier output. The data was baseline corrected based on the whole trial before implementing the independent component analysis (ICA), and down-sampled to 256 Hz for the second preprocessing step, but kept at 1024 Hz for the first. The down-sampling was done in order to decrease computational time for the classifier-locked average analyses, where the time-frequency transformation diminishes temporal resolution anyway.

Data were divided into trials from 700ms pre-stimulus to 2000ms post-stimulus onset (before implementing MVPA), or 2500ms before the classification maxima to 2500ms after the classification maxima (epochs created based on points of maximum fidelity). A high-pass filter of 0.1 Hz, a low-pass filter of 195 Hz, and a band-stop filter (48 to 52 Hz; 99 to 101 Hz, and 149 to 151 Hz), were applied to the data. At the edges of each trial, 500ms was then cut out to remove edge artifacts from filtering the epoched data. Trials were visually inspected before an ICA was computed to remove components related to eye-blink artifacts and muscle tension. After components were removed, all trials were again visually inspected, and trials

still containing artifacts were manually removed. On average 112 out of 128 trials were kept (min = 100, max = 124, SD = 7). Bad channels were interpolated using the triangulation method. Data were then re-referenced to average.

Multivariate Pattern Analysis

In order to attenuate unwanted noise, a Gaussian window with a full-width at half maximum (FWHM) in the time-domain of 40ms was applied to the signal before classification. A Linear Discriminant Analysis (LDA) was then trained and tested on the EEG sensor patterns (pre-processed signal amplitude on each of the 128 channels), independently per participant and at each time point during retrieval from 200ms pre-cue up to 1500ms post-cue. The classifier was trained to detect systematic differences between trials where participants were recalling an animate or inanimate object. A leave-one-out cross-validation procedure was used to train and test the classifier. The LDA reduces the data from 128 channels into a single decoding time course per trial, and we used these single-trial, time-resolved output of the classifier as an index of memory reinstatement. During training, the classifier found the decision boundary that could best separate the patterns of activity from the two classes (animate or inanimate) in a high-dimensional space. We then asked the classifier to estimate whether the unlabelled pattern of brain activity was more similar to one or the other class. This training-test procedure was repeated until every single retrieval trial had been classified. To avoid overfitting, the covariance matrix was regularized using shrinkage regularization [61]. The output of the classifier on a single-trial level indicates the distance to the decision boundary in a high-dimensional space, at a given time point. This parametric value is called a fidelity value or distance (d-)value, and can intuitively be regarded as reflecting how confidently the classifier predicted that the pattern of brain activity belonged to one or the other of the two classes, with the assumption being that the farther away from the boundary the more confident the classifier was [62]. Note that all the central LDA analyses in this study

were based on retrieval data. To relate retrieval phase to encoding, the same LDA approach was also applied to the encoding data. Moreover, additional results from classifiers trained on encoding and tested during retrieval are reported in the Supplemental Materials.

Power spectrum of the classifier fidelity time series

The first analysis investigated the frequency characteristics of the classifier timeseries using fast fourier transformation (FFT). This and all subsequent phase locking analyses were limited to the classifier outputs from 200ms until 1200ms after onset of the reminder. We choose this time-window of interest because based on the existing literature, memory reinstatement is highly unlikely to occur within the first 200ms post-cue, and in order to reduce influences of early, stimulus-evoked ERP components. For each participant, the trials were averaged and tapered with a Hann window before conducting the Fast Fourier Transform (FFT). To better visualize the power spectrum, a least-squares linear regression was used to subtract the 1/f background signal [63, 64]. The signal was log-transformed in the time and frequency domain and fitted with a regression line. The regression line was then subtracted from the power spectrum, and only the data that differed from the subtracted regression line were retained.

A baseline for the LDA outputs was created using a classifier with randomly shuffled labels. The labels of the two classes that the classifier later used for training and testing were shuffled pseudo randomly (to keep the same number of photographs and line drawings in each class), and fed into the LDA 25 times for each participant, such that the newly created groups had approximately the same number of trials from both classes. The parameters for running the classifier were the same as previously described for the real labels. In line with the procedure outlined in [29], and identical as for the real data, for each participant we drew (with replacement) 100 random accuracy maps (i.e., either a baseline that was created using

shuffled labels, or the real classification of the data), which were then averaged within participants. These accuracy maps were tapered with a Hann window, frequency transformed, and averaged into a group accuracy map. The background 1/f signal was subtracted using a least-squares linear regression, as described above. This procedure was repeated 1000 times, and resulted in an empirical chance distribution, which allowed us to investigate whether the results from the real-labels classification had low probability of being obtained due to chance ($p<.05$) (i.e., exceeding the 95^{th} percentile).

Phase-amplitude coupling between EEG data and fidelity values

To investigate the relationship between the continuous classifier outputs and the EEG data, the Modulation Index (MI) was computed in accordance with [65]. Following the same procedure as outlined under *Source Analysis* below, we projected the data from scalp level to source level, where each filter was computed using baseline corrected pre-processed data (-.2 – 0 sec), and frequencies below 15Hz (i.e., -200 before to 1500 ms after cue onset). Epochs were then reconstructed for 2015 virtual electrodes, rather than the original 128 electrodes. The phase of the EEG signal was estimated by convolving the data with a complex Morlet wavelet of 6 cycles. Each complex value data point was then point-wise divided by its magnitude (absolute value or complex modulus), which gave us a 4D-matrix of phase values, containing trials*channels*frequencies*time. We then binned the phase values at a given electrode (e.g. a virtual hippocampal electrode), and at a given frequency of interest (e.g. 8Hz), into 10 adjacent bins, ranging from $-\pi$ to π. The z-scored amplitudes (d-values) of the classifier output from corresponding time points were then sorted into their corresponding phase bins, and the mean amplitude of each phase bin was calculated. Following this sorting procedure at a given frequency, the modulation index was calculated. The MI was computed by comparing the distribution of classifier fidelity values across the 10 phase bins against a

uniform distribution (using the mean across bins to construct the uniform distribution). The Kullback-Leibler (KL) distance was then calculated using the equation in [65]:

$$\mathrm{D}_{KL}(P,Q) = \sum_{j=1}^{N} P(j)\log\left[\frac{P(j)}{Q(j)}\right]$$

A statistical control analysis was then performed to infer whether the MI was significantly different from a distribution that could be obtained by chance. The baseline was computed by running the same analysis as described above, but by cutting the classifier outputs into two segments at a random time point, and inserting the second data segment at the beginning of the trial. This procedure is recommended in [66], because it keeps the temporal structure of the classifier outputs largely intact while randomizing their relationship to the EEG phase at any given time point. The newly created random classifier outputs were then paired with the real EEG phase time series from their corresponding trial, and were binned in the same way as the real data. This procedure was repeated 500 times, and the MI was calculated for each iteration. The 95th percentile across iterations was determined, and the real modulation index for each subject was compared against this subject's 95th percentile using a paired samples t-test. Note that this is a very conservative analysis, resulting only in statistically significant phase modulation, if across participants real phase modulation values significantly exceed the 95th percentile of the time-permuted baseline.

Based on our initial FFT findings, all phase modulation analyses were focused on the oscillatory phase at 8Hz (Figure 2). The phase modulation index was calculated as described above for each virtual channel in source space, and a mask including left and right hippocampus (from AAL atlas as implemented in FieldTrip, see Figure 2E) was then applied to specifically extract the modulation index from our main region of interest. This was done separately for the phase modulation during retrieval, and the phase modulation during

encoding. To directly compare the preferred phase during encoding and retrieval, the bin containing the highest classifier amplitudes was identified in each participant, separately for encoding and retrieval. A Rayleigh test (implemented using *circ_rtest* in the Circular Statistics Toolbox for Matlab) was then used to statistically test the extent to which the distribution of phase angles at encoding and retrieval differed from each other.

Using classifier-locked averages to relate classifier outputs to the phase of the ongoing EEG-signal

The third, classifier-locked average analysis was aimed at characterizing the EEG phase of the time points where the classifier showed the highest fidelity. To this end, three criteria were established in order to identify times of maximum fidelity. In order to be considered a maximum, a fidelity value was required to have an amplitude that exceeded the 95th percentile of a baseline constructed from the random-label classifications. For each participant, we drew (with replacement) the fidelity timeseries from random trials 1000 times to obtain the baseline distribution. In addition, a maximum included in the final analysis was also required to remain above the 95th percentile threshold for more than 30ms, and to occur later than 200ms after reminder onset, for the same reasons as mentioned above. The average number of classification maxima extracted per trial was 2.27 (SD = 0.26). The onsets of the classifier maxima in each trial were then marked, and the corresponding time stamps were located in the raw, continuous EEG recordings. New epochs were created that were centred on each classifier maximum and contained 2.5 secs before and after the maximum, which were then cut during preprocessing to 2 secs before and after the maximum. These new epochs were used for all subsequent phase-locking analyses.

A phase-locking analysis was conducted on the new epochs to test whether classifier maxima were related to a consistent phase of a theta oscillation. For every frequency between 1 and

20 Hz, we estimated phase by convolving the data with a complex Morlet wavelet of 6 cycles. Resulting complex values were then point-wise divided by their magnitude (absolute value or complex modulus), and the mean phase was computed over all trials within each participant. The magnitude of this resulting complex value is a single value (the phase-angle time series) for each time-frequency-channel point averaged over all the trials. The value reflects the consistency of frequency-specific phase across trials and has a minimum of 0 and a maximum of 1, also called phase-locking value (PLV), phase-locking index (PLI) or Intertrial Phase Clustering (ITPC) [66].

A baseline was calculated for each trial and each participant by shifting single-trial EEG epochs randomly between 0ms and 150ms (roughly one theta-cycle) forward or backward in time, relative to the centre (i.e., the classifier maxima). By doing so, the temporal structure of the analysed signal was kept intact, but the signal was shifted relative to the classifier maxima. The phase-locking index was calculated as described above for the "real", non-shuffled data. Shuffling was done 25 times per participant and thereafter averaged together.

First, paired samples t-tests were computed between the real data and the time-shuffled baseline to investigate the difference in phase-consistency when using all maxima. To account for the multiple comparisons problem, the t-statistics for each time point (-500ms to 500ms), frequency band (6 to 14 Hz), and electrode were subjected to nonparametric cluster-based permutation testing, as implemented in the FieldTrip software. The threshold for the statistical testing was set to an alpha level of 0.025. The minimum number of neighbouring channels that were considered a cluster was set to two. T-values above the threshold of 0.1 were then summed up, and compared against a distribution where condition labels were randomly assigned 5000 times with the Monte-Carlo method, following the standard method implemented in FieldTrip.

Phase consistency is strongly biased by number of trials. For our first analysis comparing all maxima against the time-shuffled baseline, the real data and shuffled baseline contained an equal number of trials. We also ensured that all subsequent comparisons were made between conditions with exactly equal trial numbers, within each participant, including an analysis contrasting classifier maxima of high fidelity and maxima of lower fidelity, and two analyses excluding early maxima (see following two paragraphs). For the analysis contrasting conditions with high and low fidelity values, we additionally controlled the average time of the high and low classifier maxima. This was done by creating 8 time bins of equal size between 200ms and 1500ms post-cue. Fidelity values in each time bin were median split into high and low fidelity values, resulting in two matrices representing high and low fidelity trials, equally distributed across time. To calculate the phase consistency, we then followed the same procedure as described above for all maxima, except that instead of using the shuffled baseline the two groups of trials were directly compared using a non-parametric cluster-based permutation test.

To investigate the degree to which our phase-locking effects were mainly produced by classifier maxima close to the reminder word, which would be strongly influenced by the early stimulus-elicited ERP, we conducted two additional analyses excluding early classifier maxima that occurred in the first 400ms and the first 600ms post-cue, respectively, from further analysis. Otherwise, these analyses followed the same method as described for all maxima, with the same time-shuffled baseline. Similarly, an analysis using only the highest classifier maximum per trial used the same procedures and baseline described in this section for all maxima.

Event-related potential analysis

Event-related analyses were mainly conducted as sanity checks, on the one hand to investigate average signal differences between the retrieval of animate and inanimate objects locked to cue onset; and on the other hand to evaluate the average signal differences and their topography/source around the time points at which the classifier showed maximal confidence that the correct category was reinstated. For the classifier-centred analysis, we only used the 20% classifier maxima with the highest fidelity values in each of the to-be-compared classes (i.e., animate and inanimate retrieval trials), in order to enhance signal-to-noise ratio. This latter analysis included on average 48 (SD = 7.10) trials per participant. Cluster-based statistics for ERPs were conducted in the same way as for phase, except that we here focused on a narrower time window from 200ms pre- until 200ms post-maximum.

Source Analysis

A linear constrained minimum variance (lcmv) beamforming approach [33] was used to reconstruct EEG epochs in source space. The source-level results were used to obtain an approximation of the hippocampal theta phase for the phase modulation analysis, and to reconstruct classifier-locked averages (i.e., phase consistency and ERP effects) in source space [38]. Since individual MRI scans were not available, a standard MRI model was used to construct the boundary element model. The boundary element model was used in combination with individual electrode positions obtained from a Polhemus system (Colchester, Vermont, USA) to reconstruct the activity on a source level. To project the phase consistency effect from scalp level to source level, each filter was computed using frequencies below 15Hz and the entire time-window from the preprocessed data (i.e., 1500ms before to 1500ms after classifier maxima), and the original epochs were then reconstructed on 2015 virtual electrodes. Thereafter the phase-locking analysis followed the same procedure as done on scalp level. For calculating the filters for the ERP effect, we used all frequencies below 20Hz, and a time-window of 300ms pre-maxima to 300ms post-maxima. The ERP was

then calculated in the same way as on a scalp level. Note that the full-brain source reconstructions or the classifier-locked effects are only used to illustrate the most likely sources of the effects observed on scalp level (see above). We do not report additional statistics at source level, since these would be circular relative of the already known effects on scalp level. Labels of MNI coordinates were assigned based on the Lancaster, et al. [67] Talairach atlas.

Distribution of fidelity values across time

To statistically test whether the distribution of fidelity values was different from a uniform distribution across the entire retrieval time window, we manually created a uniform distribution, by producing linearly spaced values between the minimum and maximum of the real values. We then calculated the chi square statistic using the crosstab function as implemented in MATLAB, which tests whether the proportion of items in one cell is equal to the product of the proportion in that row (Figure S2A).

Time generalisation

To characterise the temporal dynamics of the classifiers, we calculated the full time generalization matrices from encoding and retrieval. These matrices show where in time classification accuracy was maximal, to which degree a classifier trained at one time point generalises to a different time point, indicating temporal stability of the underlying neural code [68]. All analyses were performed using LDA as implemented in the MVPA-Light toolbox, running on MATLAB (https://github.com/treder/MVPA-Light). Two different analyses were run: training at each time point at encoding and testing at each time point at encoding (Figure S1A); training at each time point at retrieval and testing at each time point at retrieval (Figure S1B). When analysing encoding-to-encoding generalization, data were baseline corrected (-200 to 0ms), and then z-scored per trial before running the classification.

We used a k-fold cross-validation approach with 5 folds, which was repeated twice with randomly assigned folds. When training and testing at each time point at retrieval, we did not baseline correct before the classification. However, baseline correction was applied after the classification in both analyses.

Identifying oscillating frequencies

An alternative method for detecting oscillations in time series was used in addition to our FFT approach in order to corroborate our claim that classifier outputs oscillate. This method finds time points of oscillations in the data by investigating the change in phase per unit time. We followed the method detailed in [69], with a modification for dynamic filter edges only using minimum and maximum of frequencies exceeding the 1/f distribution, made in line with [31]. Briefly, we started with raw time series data, which in our case was the z-scored fidelity values averaged within participants. Instead of creating a plateau-shaped band-pass filter based on an a priori defined frequency range, the filter was constructed based on the lowest and highest frequencies exceeding the fitted line in log-log space using robustfit in MATLAB [70]. The analytic signal was obtained by applying the Hilbert transform to the data, from where we extracted the phase angle time series. To obtain the frequency and phase at each sample, frequency sliding was applied to the data as follows: (sampling frequency*diff(unwrap(signal))/(2*π)). After this step, in order to attenuate "phase slips", we applied median filters with different length in the time domain (50ms to 400ms), wherefrom we took the median, in accordance with [69]. Frequencies that did not exceed the 1/f-fitted line were then excluded, which gave us a vector for each participant containing the frequencies and time points where an oscillation was present. We then calculated the average probability across time (200 to 1200ms post-cue, as in all other analyses using classifier output) for observing an oscillation in a given frequency between 1 and 15 Hz.

To infer whether the result that we obtained was significantly different from chance, we randomly picked one averaged random label classifier per participant. The same procedure as has been described above was applied. An average of this value was then calculated, and stored. This was done 1000 times, and resulted in an estimated chance distribution. The 95th percentile was then calculated for each frequency, and compared that to the real data (Figure 2D).

QUANTIFICATION AND STATISTICAL ANALYSIS

Behavioural data

N = 24 for all behavioural analyses.

Correlation between the two measures of remembering were highly correlated, using the Spearman's rank correlation coefficient implemented in the MATLAB function corr, and can be seen on page 4 ($r_{Spearman}$ = 0.60, p = .002).

Reaction times for the first button press when retrieving animate (Mean = 3.03 secs, SD = .95 secs, min = 1.28 secs, max = 6.01 secs) and inanimate (Mean = 2.96 secs, SD = .77 secs, min = 1.47 secs, max = 4.24 secs) objects did not differ significantly, t(1,23) = .57, p = .58. The time-window used for classification (-200ms to 1500ms around the word cue) thus only minimally overlapped with the time window where participants made a button press, and can be seen on page 4.

EEG data

N = 24 for all EEG analyses.

Power spectrum of classifier output was calculated by using the Fieldtrip function ft_freqanalysis, implemented in MATLAB. The baseline was calculated as described on page

27. Every frequency that exceeded the 95th percentile was considered significant. This was done on all 24 participants, and the results can be seen in Figure 2C.

Phase-amplitude coupling between EEG data and classifier output was calculated as described on page 6. The real data and the time-shuffled baseline were subjected to a paired-samples t-test, for hippocampal virtual channels for retrieval, $t(1,23) = 1.8191$, $p < .05$, one-sided t-test (Figure 2E), and encoding, $t(1,23) = 2.7494$, $p < .05$, one-sided t-test (Figure 2E).

All phase-consistency analyses were calculated using the following procedure. The different conditions were inserted in the Fieldtrip function ft_freqstatistics on a scalp level, and ft_sourcestatistics on a source level, implemented in MATLAB, which performs a non-parametric cluster-based permutation testing.

The p-values for the different analyses were:

For all peaks at scalp level: $p = .0002$, Figure 3B.

For all peaks at source level: $p = .0002$, Figure 3C.

High vs low fidelity trials at scalp level: $p = .003$, Figure 3D.

High vs low fidelity trials at source level: $p = .009$, Figure 3E.

ERP at scalp level: $p = .04$, Figure 4A.

ERP at source level: $p = .003$, Figure 4B.

One Peak: $p = .001$, Figure S2C.

Only correct trials: $p = .002$ and .005, Figure S2D.

Excluding 400ms: $p = .001$ and .006, Figure S2E.

Excluding 600ms: $p = .049$, Figure S2F.

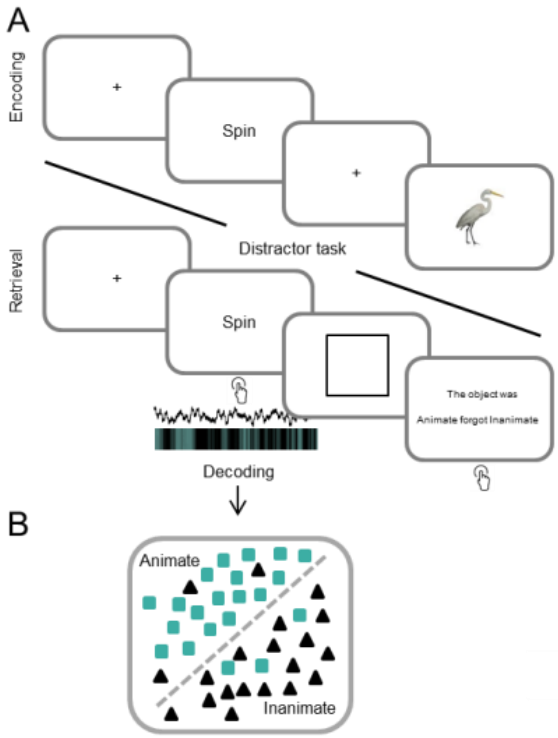
Testing for a uniform distribution, the MATLAB function crosstab was used. The function provides a chi-square test, to obtain significant difference between two distributions. The results revealed no significant difference, (χ^2 = (1, N = 6007) = 7376600, p = .375), and can be seen on page 11, Figure S2A.

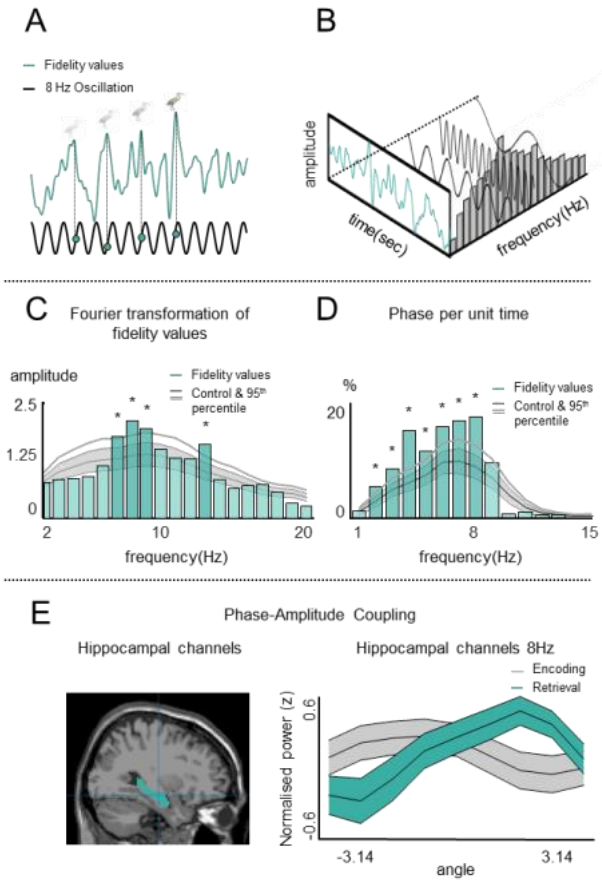
To identify oscillating frequencies, we implemented the procedure described on page 34. The results were compared to a constructed baseline, and only frequencies exceeding the 95th percentile of the baseline were considered significant (Figure 2D).

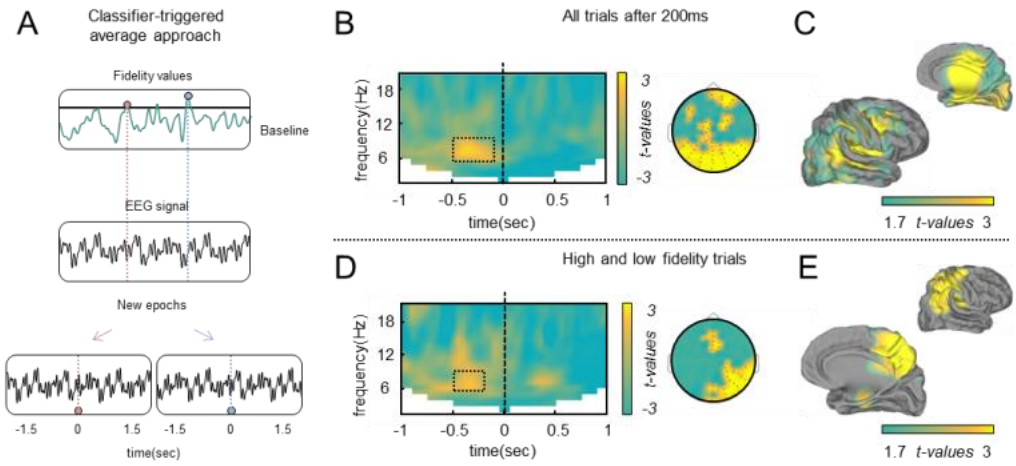
To test for difference between the power spectra for all trials and only correct trials, the two matrices were subject to a one-sided paired samples t-test, where we expected higher power for only correct trials in the 7-9Hz frequency range of interest, t(1,23) = 1.9425, p = .03 (Figure S2B).

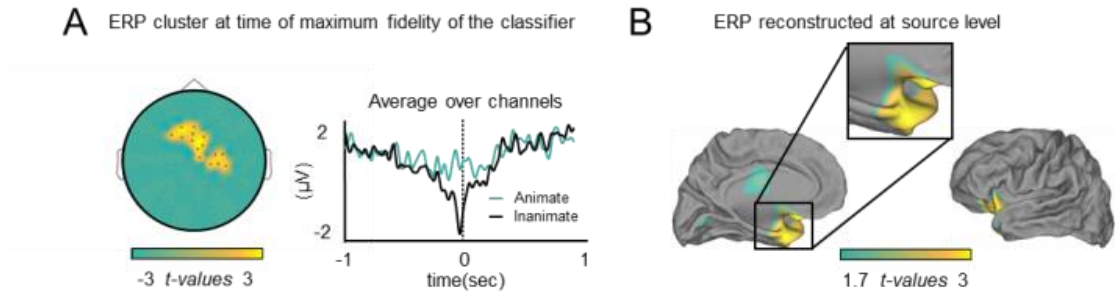
DATA AND SOFTWARE AVAILABILITY

Custom MATLAB code as well as data additional to the already published on http://dx.doi.org/10.17632/h4vcpxt4sr.1 will be made available upon request (fulfilled by Lead Contact, C.Kerren@pgr.bham.ac.uk). Since consent for sharing data at the level of the individual participant was not received originally, we can only make summary data available online or upon request.


A
Encoding
+
Spin
+
Distractor task
Retrieval
+
Spin
The object was
Animate forgot Inanimate
Decoding
B
Animate
Inanimate


A
Fidelity values
8 Hz Oscillation
B
amplitude
time(sec)
frequency(Hz)
C
Fourier transformation of fidelity values
amplitude
Fidelity values
Control & 95th percentile
2.5
1.25
0
2
10
20
frequency(Hz)
D
Phase per unit time
%
Fidelity values
Control & 95th percentile
20
0
1
8
15
frequency(Hz)
E
Phase-Amplitude Coupling
Hippocampal channels
Hippocampal channels 8Hz
Encoding
Retrieval
Normalised power (z)
0.6
-0.6
-3.14
3.14
angle


A
Classifier-triggered
average approach
Fidelity values
Baseline
EEG signal
New epochs
-1.5
0
1.5
-1.5
0
1.5
time(sec)
B
All trials after 200ms
frequency(Hz)
18
12
6
-1
-0.5
0
0.5
1
time(sec)
3
t-values
-3
C
1.7 t-values 3
D
High and low fidelity trials
frequency(Hz)
18
12
6
-1
-0.5
0
0.5
1
time(sec)
3
t-values
-3
E
1.7 t-values 3


A
ERP cluster at time of maximum fidelity of the classifier
Average over channels
2
-2
(μV)
Animate
Inanimate
-1
0
1
time(sec)
-3 t-values 3
B
ERP reconstructed at source level
1.7 t-values 3

Supplemental Figures

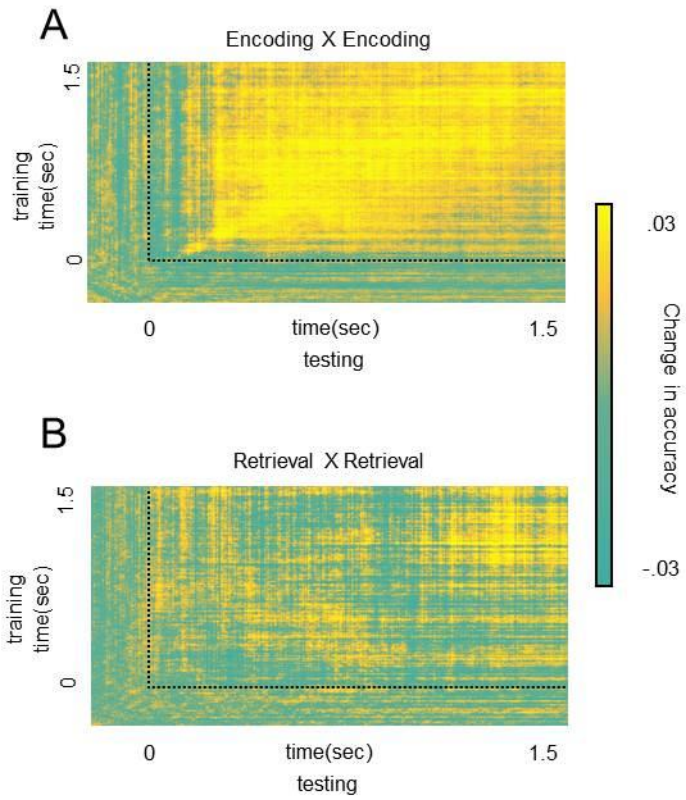


Figure S1. Time generalisation matrices for encoding and retrieval, with time zero indicating the onset of the object during encoding, and the onset of the reminder word during retrieval, related to STAR Methods.
(A) Training and testing at encoding showed sustained high classifier accuracy from approximately 500-600ms to the end of the time window. **(B)** Training and testing at retrieval shows that accuracy is generally above baseline after cue onset, and indicates that participants reinstated the memory at different time points, and possibly several times. Unlike at encoding, the retrieval pattern suggests that there is not a sustained state across the entire time period, consistent with periodic reactivation. Each of the matrices in panels A-B is based on an LDA classification using a 5-fold cross-validation.

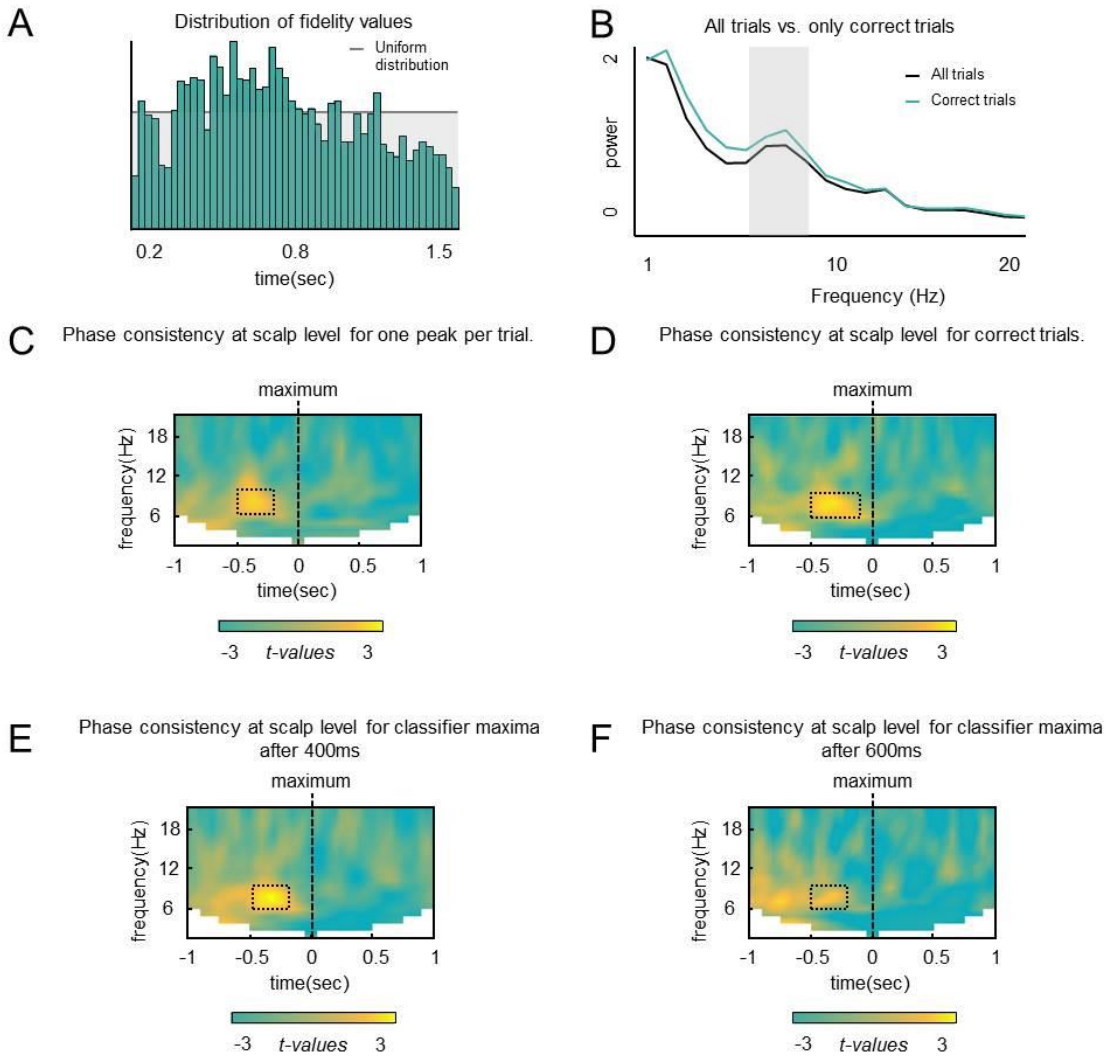


Figure S2. Distribution of classifier maxima across participants and time, behavioural relationship to power spectra, and phase consistency for various control analyses, related to Figure 3 and STAR Methods.
(A) The distribution of classifier maxima, accumulated across participants, showed no significant deviation from a uniform distribution, indicating that the maxima were evenly distributed across the entire retrieval period, with a noticeably increased density around 400-800ms. This is in line with previous studies showing strongest memory reinstatement in the recollection period. **(B)** To evaluate the relationship between the power spectra and memory performance, we compared the power spectra for all trials and correct trials only for 7-9 Hz, which revealed a significantly stronger effect for correct trials compared to all trials. Note that a direct comparison between correct and incorrect trials was not possible due to a low number of incorrect trials in the cued recall task. **(C)** Classifier-locked averages showing phase consistency when using only the highest maximum per trial, and thus excluding all overlapping epochs. As expected, the phase consistency is less temporally smeared, with a cluster from -500ms to -150ms pre-maximum. **(D)** Same analysis as shown in main Figure 3, but limited to correct trials, showing a cluster of significant phase consistency 500-150ms before the classifier maxima. **(E)** Removing the first 400 ms of classifier maxima did not change the phase consistency effect, neither did removing the first 600 ms of classifier maxima **(F)**. When using only very late maxima, an even earlier cluster of 7-8Hz phase consistency becomes evident, with the later cluster at -500ms to -250 ms remaining significant. This results likely reflects several cycles of a 7-8Hz oscillation.

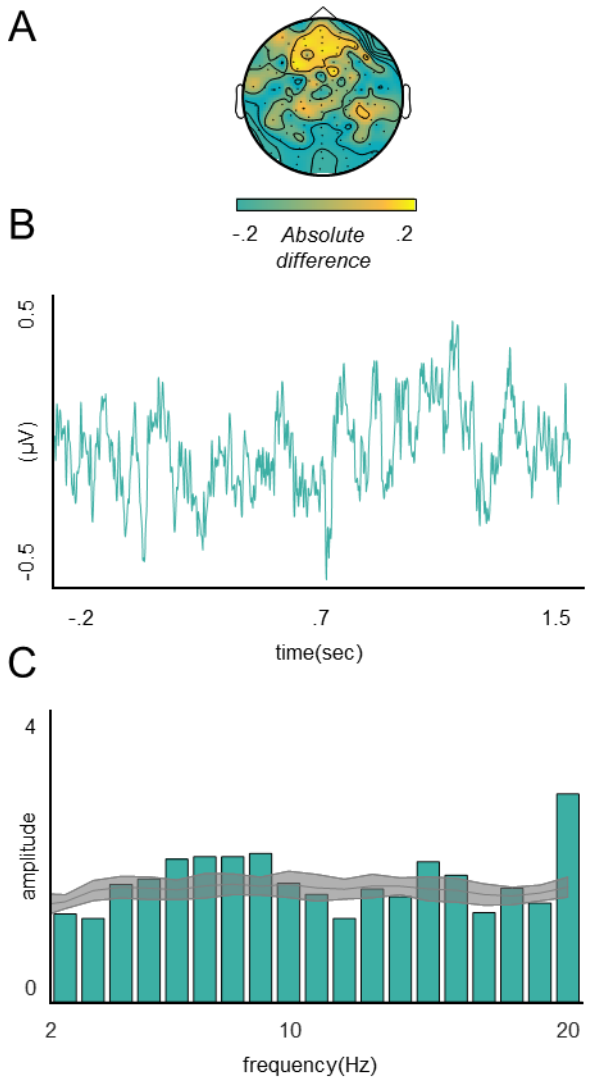


Figure S3. Average difference between animate and inanimate object retrievals shown in the time and frequency domain, related to Figure 4 and STAR Methods.
(A) Topographies of the absolute EEG difference between the recall of animate and inanimate objects, showing a frontal maximum during retrieval (600-1200ms). **(B)** Average difference signal between animate and inanimate objects during retrieval, interestingly showing a visible rhythmicity. **(C)** Applying the Fourier-transform, we can see above baseline power increases in spectral frequencies between 6-9Hz, the same frequencies that also show power increases in the classifier time series.

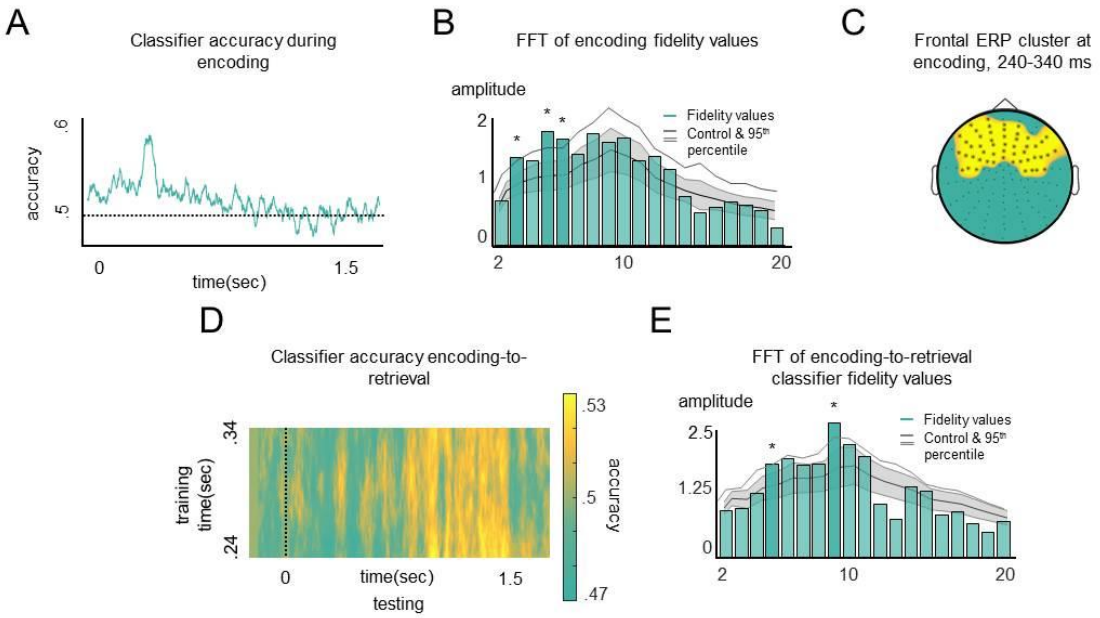


Figure S4. Encoding for encoding and encoding for retrieval analyses, related to Figure 2 and STAR Methods.
(A) Training and testing at encoding revealed a peak of classifier performance at ~300ms, a time window commonly seen when investigating encoding activity for semantic memory. **(B)** The averaged fidelity values were subjected to a Fourier Transformation, and showed a peak in the lower frequencies. **(C)** At encoding, a frontal cluster survived a non-parametric cluster-based permutation test, indicating an overlap with retrieval activity seen in Figure S3A. **(D)** Using the time points from 240-340ms during encoding, where animate vs inanimate differences showed an overlapping topography compared with retrieval (see Fig. S4A), a fluctuating pattern is also visible in the time generalisation matrix of a classifier trained on encoding and tested during retrieval. **(E)** The power spectra of this encoding-retrieval classifier revealed a peak at 9 Hz.